MEME16203Linear Models

Assignment 5

UNIVERSITI TUNKU ABDUL RAHMAN

Faculty:	FES	Unit Code:	MEME16203
Course:	MAC	Unit Title:	Linear Models
Year:	1,2	Lecturer:	Dr Yong Chin Khian
Session:	May 2024		
Due by:	9/8/2024		

Q1. A study was conducted on human subjects to measure the effects of 3 foods on serum glucose levels. Each of the 3 foods was randomly assigned to 4 subjects. The serum glucose was measured for each of the subjects at 5 different time points starting at 15 minutes and every 15 minutes after food was ingested. Consider the model

$$Y_{ijk} = \mu + \alpha_i + S_{ij} + \tau_k + \gamma_{ik} + e_{ijk}$$

where y_{ijk} is the serum glucose levels at the k^{th} time point for the j^{th} subject with the i^{th} food, α_i is the fixed diet effect, τ_k is the fixed time effect and γ_{ik} is the fixed diet $\times$ time effect, $S_{ij} \sim NID(0, \sigma_S^2)$ and is independent of $e_{ijk} \sim NID(0, \sigma_e^2)$.

(a) Find $V(\mathbf{Y_{ij}})$, for this model? (5 marks)

(b) Provide the formulas for the estimator of σ_e^2 and σ_S^2. (5 marks)

(c) What is the correlation between observations taken on the same subject? (5 marks)

(d) Find the estimator of $V(\bar{Y}_{ij.})$ and provide it's degrees of freedom. (5 marks)

(e) Find the estimator of $V(\bar{Y}_{i.k})$ and provide it's Satterthwaith degrees of freedom. (5 marks)

[Total:25 marks]

Q2. An experiment was conducted to study the amount of dry matter measured for wheat plants grown under conditions with different levels of moisture and different amounts of fertilizer. The experiment was conducted in 4 greenhouses. Each greenhouse contained four tables. On each table, there were 4 pots with 1 wheat seed planting in each pot. Before planting the wheat seeds, fertilizer was added to the soil in the pots at levels of 2, 4, 6, or 8 mg. per pot. The four levels of fertilizer were randomly assigned to the four pots within each table. The levels of the moisture factor corresponded to adding either 10, 20, 30, or 40 ml. of water per pot per day to the table. The water was absorbed thorough holes in the bottom of the pots. Moisture levels were randomly assigned to tables with one table assigned to one moisture level. Thus, the four plants on any given table received the same amount of water throughout the experiment. At the conclusion of the experiment, the amount of dry matter of each plant was recorded.

For $i = 1,2,3,4; j = 1,2,3,4.$ and $k = 1,2,3,4$; let y_{ijk} denote the observed dry matter weight for the wheat grown in the pot associated with greenhouse i, watering level j, and moisture k.

$$y_{ijk} = \mu + g_i + \omega_j + t_{ij} + \gamma_k + \phi_{jk} + e_{ijk} - - - - - - -(1)$$

where the g_i terms are $N(0, \sigma_g^2)$, the t_{ij} terms are $N(0, \sigma_t^2)$, the e_{ijk} terms are $N(, \sigma_e^2)$, all these random terms are mutually independent, and the remaining terms in the model are unknown fixed parameters.

(a) According to MODEL 1, what is the correlation between the heights of two plants growing together on the same table? (5 marks)

(b) In terms of the MODEL 1 parameters, write down the null hypothesis of no watering level main effects. (5 marks)

(c) MODEL 1 can be written as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, where

$$\mathbf{y} = (y_{111}, y_{112}, y_{113}, y_{114}, y_{121}, y_{122}, y_{123}, y_{124}, \ldots, y_{441}, y_{442}, y_{443}, y_{444}).$$

Provide corresponding expressions for $\mathbf{X}$, $\boldsymbol{\beta}$, $\mathbf{Z}$, and $\mathbf{u}$. Express $\mathbf{X}$ and $\mathbf{Z}$ using Kronecker product notation. (5 marks)

(d) Report the approprite degress of freedom, expected mean squares in the ANOVA table and use the results to obtain formulas for estimates of the variance components. (5 marks)

(e) Find the estimator of $V(\bar{Y}_{i..})$ and provide it's Satterthwaith degrees of freedom. (5 marks)

[Total:25 marks]

Q3. An experiment was conducted to compare the effectiveness of two diets (denoted D_1 and D_2). The subjects included 60 men between the ages of 18 to 35. Each subject lifts the dumbbells until his muscles were depleted of energy, rested for two hours, and lifts the dumbbells again until exhaustion. During the rest period, each subject eat one of the two diets as assigned by the researchers. Each subject's performance on the second round of lifting dumbbells following the rest period was assigned a score between 0 and 100 based on the energy expended prior to exhaustion. Higher scores indicate of better performance. 20 of the 60 subjects repeated the lift rest lift trial on a second occasion separated from the first by approximately three weeks. These subjects eat one diet during the first trial and the other during the second trial. The diet order was randomized for each subject by the researchers. The other 40 subjects performed the trial only a single time, eating a randomly assigned diet during the rest period. 20.0 of these subjects received diet 1, and the other 20.0 received diet 2. A portion of the entire data set is provided in the following table.

Subject	Diet 1	Diet 2
1	40	50
2	35	38
⋮	⋮	⋮
20	28	55
21	35	–
22	80	-
⋮	⋮	⋮
40.0	55	-
41.0	-	17
42.0	-	54
⋮	⋮	⋮
60	-	60

Subjects 1 through 20 in the table above represent the 20 subjects who performed the trial separately for each of the diets. Note that the data set contains no information about which diet was received in the first trial and which drink was received in the second trial. Suppose the following model is appropriate for the data.

$$y_{ij} = \mu_i + u_j + e_{ij}, (1)$$

where y_{ij} is the score for diet i and subject j μ_i is the unknown mean score for diet i, u_j is a random effect corresponding to subject j, and e_{ij} is a random error corresponding to the score for diet i and subject j ($i = 1, 2$ and $j = 1, 2, \ldots, 60$). Here $u_1, u_2, \ldots, u_{60}$ are assumed to be independent and identically distributed as $N(0, \sigma_u^2)$ and independent of the e_{ij}'s, which are assumed to be independent and identically distributed as $N(0, \sigma_e^2)$.

(a) For each of the subjects who received both diets, the difference between the scores ($D_1 - D_2$ score) was computed. This yielded 20 score differences denoted $d_1, d_2, \ldots, d_{20}$. Describe the distribution of these differences considering the assumptions about the distribution of the original scores in model (1). (9 marks)

(b) Suppose you were given only the differences from part (a). Provide a formula for a test statistic (as a function of $d_1, d_2, \ldots, d_{20}$) that could be used to test $H_0 : \mu_1 - \mu_2$. (8 marks)

(c) Let $a_1, a_2, \ldots, a_{20.0}$ be the scores of the subjects who received only diet 1. Let $b_1, b_2, \ldots, b_{20.0}$ be the scores of the subjects who received only diet 2. Suppose you were given only these 20.0 scores. Determine the BLUE of $\mu_1 - \mu_2$ and the variance of this estimator. (8 marks)

[Total:25 marks]

Q4. Researchers investigated the effects of 18 different drugs on the level of a protein in the blood of mice. On each of 8 days, 18 mice were randomly assigned to the 18 drugs with one mouse for each drug. Each mouse was injected with its assigned drug, and then blood samples were taken from each mouse at 4 time points: 1, 2, 3, and 4 hours after injection. The same process was repeated each day with 18 different mice, so a total of 144 mice were used in the experiment. The level of the protein of interest was measured in each of the 576 blood samples. For $i = 1, \ldots, 8$, $j = 1, \ldots, 18$, and $k = 1, \ldots, 4$, let y_{ijk} be the protein level measurement on day i for chemical drug j at time k. For $i = 1, \ldots, 8$, $j = 1, \ldots, 18$, and $k = 1, \ldots, 4$, consider the model

$$y_{ijk} = \mu_{jk} + d_i + e_{ijk},$$

where μ_{jk} terms are unknown fixed parameters and the other terms are random effects defined as follows. Let $\mathbf{d} = [d_1, \ldots, d_8]^T$. For $i = 1, \ldots, 8$ and $j = 1, \ldots, 18$, let $\mathbf{e}_{ij} = [e_{ij1}, \ldots, e_{ij4}]^T$.

Suppose

$$\mathbf{d} \sim N(\mathbf{0}, \sigma_d^2 \mathbf{I}_{8\times 8}),$$

and

$$\mathbf{e}_{ij} \sim N(\mathbf{0}, \mathbf{\Sigma}_e) \text{ for } i = 1, \ldots, 8 \text{ and } j = 1, \ldots, 18,$$

where σ_d^2 is an unknown positive variance parameter and

$$\mathbf{\Sigma}_e = \begin{bmatrix} \sigma_1^2 & \rho^4\sigma_1\sigma_2 & \rho^9\sigma_1\sigma_3 & \rho^{19}\sigma_1\sigma_4 \\ \rho^4\sigma_1\sigma_2 & \sigma_2^2 & \rho^5\sigma_2\sigma_3 & \rho^{15}\sigma_2\sigma_4 \\ \rho^9\sigma_1\sigma_3 & \rho^5\sigma_2\sigma_3 & \sigma_3^2 & \rho^{10}\sigma_3\sigma_4 \\ \rho^{19}\sigma_1\sigma_4 & \rho^{15}\sigma_2\sigma_4 & \rho^{10}\sigma_3\sigma_4 & \sigma_4^2 \end{bmatrix}$$

for some unknown standard deviation parameter $\sigma_i > 0, i = 1, 2, 3, 4$ and some unknown correlation parameter ρ. Finally, suppose that $\mathbf{d}$ and $\mathbf{e}_{11}, \ldots, \mathbf{e}_{8,18}$ are all independent. In terms of model parameters, give a simplified expression for the variance of the generalized least squares estimator of each of the following:

(a) μ_{34} (6 marks)

(b) $\bar{\mu}_{3.}$ (7 marks)

(c) $\mu_{13} - \mu_{23}$ (6 marks)

(d) $\mu_{31} - \mu_{34}$ (6 marks)

[Total:25 marks]